Short communication

Estimation of stream channel geometry in Idaho using GIS-derived watershed characteristics

Daniel P. Ames a,⁎, Eric B. Rafn a, Robert Van Kirk b, Benjamin Crosby c

a Idaho State University, Geosciences, 1776 Science Center Dr. Ste 115, Idaho Falls, ID 83402, USA
b Idaho State University, 921 S. 8th Ave., Stop 8085, Pocatello, ID 83209-8085, USA
c Idaho State University, 521 S. 8th Ave., Stop 8072, Pocatello, ID 83209-8072, USA

A R T I C L E   I N F O

Article history:
Received 23 May 2008
Received in revised form 1 August 2008
Accepted 14 August 2008
Available online 5 November 2008

Keywords:
Stream channels
Modeling
Watersheds
GIS
Streamstats
Cross-sections

A B S T R A C T

This paper describes estimation of stream channel geometry with multiple regression analysis of GIS-derived watershed characteristics including drainage area, catchment-averaged precipitation, mean watershed slope, elevation, forest cover, percent area with slopes greater than 30 percent, and percent area with north-facing slopes greater than 30 percent. Results from this multivariate predictor method were compared to results from the traditional single-variable (drainage area) relationship for a sample of 98 unregulated and undiverted streams in Idaho. Root-mean-squared error (RMSE) was calculated for both multiple- and single-variable predictions for 100 independent, random subsamples of the dataset at each of four different subsample levels. The multiple-variable technique produced significantly lower RMSE for prediction of both stream width and depth when compared to the drainage area-only technique. In the best predictive equation, stream width depended positively on drainage area and mean watershed precipitation, and negatively on fraction of watershed consisting of north-facing slopes greater than 30%. Stream depth depended positively on drainage area and precipitation, and negatively on mean watershed elevation. Our results suggest that within a given physiographic province, multivariate analysis of readily available GIS-derived watershed variables can significantly improve estimates of stream width and depth for use in flow-routing software models.

⁎ Corresponding author. Fax: +1 208 282 7716.
E-mail address: amesdan@isu.edu (D.P. Ames).

1. Introduction

Streamflow simulation models such as Hydrologic Simulation Program – Fortran (HSPF) (Bicknell et al., 2001), Soil and Water Assessment Tool (SWAT) (Arnold et al., 1993, 1998), Agricultural Nonpoint Source Pollution Model (AGNPS) (Young et al., 1987, 1989), and Water Analysis Simulation Program version 4 (WASP4) (Ambrose et al., 1988); erosion and sediment transport models such as those reviewed in Merritt et al. (2003); and various other catchment scale environmental models (e.g. Croton and Barry, 2001; Sandersona and Baginska, 2006; Katiyara and Hossain, 2007), typically require stream channel geomorphologic data as input parameters (e.g. channel length, slope, average width, depth, and flood-plain size). Spatio-temporal deterministic models simulate hydrologic processes in a watershed by routing streamflow through virtual channels using equations for mean flow velocity (e.g. Manning’s equation). Manning’s equation is based on channel roughness, slope and hydraulic radius, R (Manning, 1851). This last parameter, R, is typically calculated by summing the channel width with two times the channel depth. Hence in the absence of full geomorphological data, average stream width and depth are required at a minimum as model parameters. This paper presents a novel, GIS-based technique for estimating channel width and depth for streamflow simulation models.

Acquiring complete channel geometry measurements can be time-consuming and cost-prohibitive (Lacroix et al., 2002), motivating the current interest in obtaining estimated channel geometry parameters directly from readily available GIS data. Whereas channel length and slope can be measured using digital elevation models (DEMs) and channel extraction algorithms, average width and depth are not as easily estimated. These parameters are typically approximated as having a power-law dependence on river discharge (e.g. Leopold and Maddock, 1953). Though this method has served hydrologists and geomorphologists well for the last 50 years, it limits analysis to stream reaches with known established discharge records. To extend analysis to catchments without discharge records, many have employed another power-law relation between river discharge and upstream drainage area to approximate channel geometry from DEMs alone. The weakness in this method is that it assumes that discharge increases systematically with increasing drainage area. The assumption that discharge scales systematically with drainage area fails to recognize how
spatial variations in topography, land cover, and climate directly influence discharge. Further, currently used equations (equations (1) and (2)), for predicting width and depth from drainage area were obtained from large-scale studies of watersheds across multiple physiographic provinces in the United States and may not be applicable to small-scale watershed studies. This paper evaluates whether the inclusion of multiple GIS-derived spatial variables into stream channel geometry estimation equations improves regional predictions of width and depth.

The hydrologic simulation models cited previously all use simplified representations of stream channel cross-sections such as nested trapezoids, a single trapezoid, or a rectangular cross-section. Our goal is to estimate average depth and bankfull width for use in rectangular cross-section simplification of the channel. An optimal stream channel description would include measurements of how width varies with increasing depth; however because most hydrologic models are not designed to process complex channel geometry data, a simplified GIS-based estimation procedure is justified.

The US Environmental Protection Agency (EPA) Better Assessment Science Integrating Point and Nonpoint Sources (BASINS) watershed analysis system (Battin et al., 1998) currently includes functions for estimating an average bankfull width and depth using a methodology based upon Muttiah et al. (1997), Allen et al. (1994), and Leopold and Maddock (1953). The equations used in the BASINS software to estimate stream width and depth, respectively, are

\[ W = 1.29A^{0.66} \]  

and

\[ D = 0.13A^{0.4} \]  

where \( W \) = bankfull channel width (m), \( D \) = bankfull channel depth (m), and \( A \) = drainage area (km\(^2\)).

The advantages of equations (1) and (2) are that the studies from which they are derived include multiple physiographic provinces in the United States, and can serve as consistent average width and depth predictors when applied nationally. However, they cannot account for watershed parameters, such as precipitation, that vary from one physiographic province to another. These parameters directly affect width and depth predictions. For example, a low relief desert watershed and a forested, mountainous coastal watershed of similar drainage area are not likely to have similar channel geometries.

Muttiah et al. (1997) and Berenbrock (2002) demonstrated that streamflow can be predicted from multiple physical watershed characteristics with improved predictions for both large and small-scale watersheds. We hypothesize that this approach can be applied to improve stream width and depth estimates for unregulated and undiverted streams in Idaho. In this study, improved prediction is defined by smaller root-mean-squared error (RMSE) between predicted and observed values. Studies are unregulated and undiverted streams in and adjacent to Idaho with gage stations operated by the US Geological Survey (USGS). Datasets analyzed include field measurements of bankfull width and depth at these gage stations as well as basin characteristics collected by the USGS and published by Berenbrock (2002).

2. Methods

Watershed and stream characteristics were collected for Idaho and neighboring states (some watersheds in this dataset extend into neighboring states). To assure that channel geometry estimates are derived primarily from natural watershed characteristics and not water management, only unregulated and undiverted streams were used in this study. Stream channel characteristics, including bankfull stream width and depth at each gage station, were collected from the online USGS National Water Information System (NWIS) database (USGS, 2007). These data were recorded periodically by USGS field workers and included wetted widths and cross-sectional areas at each location. For this study, only widths and depths associated with two-year peak streamflow events were used.

Formal characterization of bankfull streamflow includes analysis of the stage-discharge relationship at each gage station, taking into consideration actual stream channel geometry. However, Wolman and Miller (1960) identified the two-year peak stream discharge as the “dominant discharge” considered to do the most work in establishing channel geometry. Hence, we used the two-year peak stream discharge as the representative dominant discharge flow regime at each gage station for which width and depth measurements were collected. Two-year peak stream discharge was estimated using the median annual peak flow from all available years within the USGS NWIS datasets. At the median value 1/2 of the data are higher (and lower), hence it represents a two-year return. Although 137 Idaho USGS gage stations were identified as unregulated and undiverted and had available width and depth measurements, these measurements were recorded at a variety of flow regimes. Gage station selection was limited to field records that were collected at flows within one standard deviation of the estimated two-year peak stream discharge. Of the 137 gage stations, 93 met this flow criterion and thus define our sample (Fig. 1). Drainage area for these 98 gage locations ranged from 1 to 8,400 km\(^2\).

We used least-squares linear regression analysis to investigate the dependence of stream width and depth, respectively, on watershed characteristics. Berenbrock (2002) had previously compiled watershed characteristic data for the sample gage stations used in our study. From these data, the watershed characteristics used as explanatory variables in our regression analyses were drainage area (\( A \)), mean watershed elevation (\( E \)), mean annual precipitation (\( P \)), forested area (\( F \)), mean watershed slope (\( M \)), fraction of watershed area with slopes greater than 30 percent (\( S \)), and fraction of watershed area consisting of north-facing slopes greater than 30 percent (\( N \)). Lithology was not used in the study because previous research indicated that lithology was less influential than other watershed characteristics in predicting stream morphology in Idaho (Bayrd, 2006).

Preliminary analysis of regression relationships and residuals showed significant curvature from required assumptions for several of the variables. The stream characteristics width (\( W \)) and depth (\( D \)) and basin characteristics \( A \) and \( P \) were logarithmically transformed for use in the analysis to meet the assumptions of linearity in variable relationships and homoscedasticity in residuals, and to prevent data from a few large basins exerting undue influence over regression equations (Neter et al., 1989; Helsel and Hirsch, 1992; Ramsey and Schafer, 2002).

Prior to conducting the regression analysis, we eliminated highly correlated independent variables. Pearson correlation coefficients exceeded 0.95 for all pairwise combinations of \( S, N, A \), and \( M \). Preliminary multiple regression analysis showed that stream width and depth depended more on percent of watershed with north-facing slopes greater than 30% than on the other two slope variables, so we chose \( N \) as the representative slope variable. Mean watershed precipitation, \( P \) (in-transformed) and percent of watershed with forest cover (\( F \)) were also highly correlated (\( r = 0.82 \)). Preliminary analysis showed that stream width and depth depended much more on precipitation, so we eliminated percent forest cover. Thus, the four watershed variables used in the regression analysis were \( A, \ln(P), \) and \( N \). Pearson correlations among these four variables were less than 0.37.

We used bootstrap sampling to test the hypothesis that a multivariate regression model better estimates stream width and depth than a single-variable (drainage area-only) model. We randomly selected a subsample of 78 of the \( n = 98 \) streams (\( p = 80\% \) of the sample), using a random multivariate regression analysis to determine if this reduced sample size, \( \frac{N}{2} \), represented a significant decrease in the sample size. However, the results were not significant for any of the remaining \( 10 \) subsamples. Therefore, we conclude that a multivariate regression model that uses only two variables is not significantly different than a model using all four variables.

We computed the root-mean-squared error (RMSE) between predicted widths and depths over the remaining 20 streams \( (1 - p = 20\%) \). The root-mean-squared error (RMSE) between observed and predicted widths and depths over the remaining 20 streams was then calculated from each regression equation. This process was performed on each of 100 independent subsamples of size \( 78 \), yielding bootstrap samples of 80 multivariate regression equations. The bootstrap procedure was repeated for 100 independent subsamples comprising, respectively, \( p = 60\%, p = 40\% \), and \( p = 20\% \) of the 98-stream sample (RMSE computed from the remaining 40%, 60%, and 80%, respectively).

For each of the four subsample levels, the performance of the regression model (multivariate versus drainage area-only) was assessed by comparing RMSEs for the full regression method with that for the reduced regression method with a paired \( t \)-test. Hypothesis tests were performed at the \( \alpha = 0.05 \) significance level with the one-sided alternative hypothesis that RMSE was smaller for the multivariate method. After using the bootstrap tests to assess which method produced the best predictive equation, we applied the best method to the entire sample of 98 gage stations to determine field predictive equations for width and depth across the entire study area.

3. Results and discussion

Mean RMSE on stream width prediction for the multivariate method was significantly lower than that for the drainage area-only
method at all subsample levels (Table 1, Fig. 2). Mean RMSE on stream depth prediction for the multivariate method was significantly lower than that for the drainage area-only method at all but the 20% subsample level (Table 2, Fig. 3). These hypothesis test results show that the multivariate method better predicts stream width even when a relatively small sample of stream gage sites is used to determine the predictive equations. The multivariate method also better predicts stream depth, but the reduction in error may not be statistically significant if the sample from which the regression equation is determined is relatively small. Because stream width shows more variation among streams of different watershed characteristics than does stream depth (coefficients of variation for ln-transformed stream width and depth in our dataset were 0.30 and 0.10, respectively), error between observed and predicted stream widths will be smaller when more explanatory variables are included, even at small sample sizes. Somewhat larger sample sizes are required to realize this error reduction in stream depth prediction. For predictions of stream width and stream depth, RMSE for both multivariate and single-variable methods

<table>
<thead>
<tr>
<th>p</th>
<th>Mean RMSE: multivariate method</th>
<th>Mean RMSE: drainage area-only method</th>
<th>Mean paired RMSE difference</th>
<th>DF</th>
<th>t-test, P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>15.29</td>
<td>23.43</td>
<td>-8.14</td>
<td>99</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>0.4</td>
<td>13.84</td>
<td>21.99</td>
<td>-8.15</td>
<td>99</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>0.6</td>
<td>13.61</td>
<td>21.45</td>
<td>-7.84</td>
<td>99</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>0.8</td>
<td>13.02</td>
<td>20.36</td>
<td>-7.34</td>
<td>99</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
were annual watershed precipitation (m), north-facing slopes with gradients greater than 30%, and depth at the two-year flood (m).

Mean RMSE: $t$-test, $P$-value

<table>
<thead>
<tr>
<th>$p$</th>
<th>Mean RMSE: multivariate method</th>
<th>Mean RMSE: drainage area-only method</th>
<th>Mean paired RMSE difference</th>
<th>DF</th>
<th>$P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.7946</td>
<td>0.8000</td>
<td>-0.00522</td>
<td>99</td>
<td>0.351</td>
</tr>
<tr>
<td>0.4</td>
<td>0.7162</td>
<td>0.7694</td>
<td>-0.0531</td>
<td>99</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>0.6</td>
<td>0.6750</td>
<td>0.7496</td>
<td>-0.0746</td>
<td>99</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>0.8</td>
<td>0.6846</td>
<td>0.7397</td>
<td>-0.05522</td>
<td>99</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Results of paired $t$-tests comparing RMSE for predicted stream depth between multivariate and single-variable (drainage area) methods. The value of $p$ is the fraction of the dataset used to determine the prediction equations. DF stands for degrees of freedom. All tests used the one-sided alternative hypothesis that RMSE for the multivariate method was less than that for the single-variable method.

Results of the stepwise multiple regression analysis for stream width applied to the entire dataset (98 stream gage sites). MSE stands for mean squared error, AIC represents Akaike’s Information Criterion.

<table>
<thead>
<tr>
<th>Variables in model</th>
<th>Adj. $R^2$</th>
<th>MSE</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A, P, N, E$</td>
<td>0.869</td>
<td>0.130</td>
<td>-192.2</td>
</tr>
<tr>
<td>$A, P, N$</td>
<td>0.867</td>
<td>0.131</td>
<td>-193.2</td>
</tr>
<tr>
<td>$A, P$</td>
<td>0.862</td>
<td>0.136</td>
<td>-191.5</td>
</tr>
<tr>
<td>$A$</td>
<td>0.649</td>
<td>0.347</td>
<td>-101.7</td>
</tr>
</tbody>
</table>

Results of the stepwise multiple regression analysis for stream depth applied to the entire dataset (98 stream gage sites). MSE stands for mean squared error, AIC represents Akaike’s Information Criterion.

<table>
<thead>
<tr>
<th>Variables in model</th>
<th>Adj. $R^2$</th>
<th>MSE</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D, P, N$</td>
<td>0.773</td>
<td>0.147</td>
<td>-180.0</td>
</tr>
<tr>
<td>$D, P$</td>
<td>0.776</td>
<td>0.145</td>
<td>-183.0</td>
</tr>
<tr>
<td>$D$</td>
<td>0.709</td>
<td>0.150</td>
<td>-181.9</td>
</tr>
<tr>
<td>$D$</td>
<td>0.646</td>
<td>0.229</td>
<td>-142.3</td>
</tr>
</tbody>
</table>

Fig. 3. RMSE for stream depth predicted by regression equations derived, respectively, using multiple explanatory variables and a single variable (drainage area). The value of $p$ is the fraction of the dataset used to determine the prediction equations. Each box plot summarizes values for 100 independent random samples.
4. Conclusions

Stream channel characteristics, including average stream width and depth, are important parameters for hydrologic model users to help define flow routing through a simulated watershed. Many current watershed delineation tools, such as BASINS, that estimate average width and depth use only drainage area. This study shows that the use of regional data and additional watershed variables can result in better regression equations to predict stream width and depth. The variables with the greatest predictive power in Idaho were drainage area and precipitation. Slope and elevation variables had secondary effects on prediction of stream width and depth.

These results indicate that estimated stream channel geometry from watershed characteristic data will result in better flow routing, aiding hydrologic simulation model developers and users. The approach presented here should improve parameterization of streamflow models by providing a more accurate method for channel geometry estimation as compared to other existing methods. In the United States, watershed data are readily accessible for several states and can be easily calculated from digital geospatial data layers using standard GIS tools for other locations. Future work might include the development of modeling support tools to automate parameter selection and coefficient determination for equations for a specific region of interest.

Acknowledgements

This work was supported in part by the NSF-Idaho EPSCoR Program and by the National Science Foundation under award number EPS-0447689. The authors express gratitude to Raghavan Srinivasan, Charles Berenbrock, and Scott N. Miller for providing insight and guidance as well as Kathleen Kakeacek, Teri Peterson, Bryce Contor, Maxine Dakins and Trish Mercer for their technical help with the manuscript.

References